UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/642,452	08/18/2000	Josef Bauer	POO,1701	7124

7590 03/16/2004

Morrison & Foerster LLP
1650 Tysons Boulevard
Suite 300
McLean, VA 22102

EXAMINER	
LERNER, MARTIN	

ART UNIT	PAPER NUMBER
2654	16

DATE MAILED: 03/16/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/642,452	BAUER ET AL.	
	Examiner	Art Unit	
	Martin Lerner	2654	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *26 November 2003 and 15 January 2004*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *15 to 28* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *15 to 28* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☒ All b)☐ Some * c)☐ None of:

1.☒ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

Drawings

1. New formal drawings are required incorporating the proposed drawing changes submitted 26 November 2003, which changes are approved. The corrected drawings are required in reply to the Office action to avoid abandonment of the application. The requirement for corrected drawings will not be held in abeyance.

Claim Objections

2. Claims 15 to 27 are objected to because of the following informalities:

In claim 15, line 7, the term "volume distance" lacks antecedent basis. Presumably, "volume distance" should be –volume difference— as the preceding clause sets forth the step of calculating a difference between volumes. Appropriate correction is required.

Claim Rejections - 35 USC § 102

3. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (e) the invention was described in a patent granted on an application for patent by another filed in the United States before the invention thereof by the applicant for patent, or on an international application by another who has fulfilled the requirements of paragraphs (1), (2), and (4) of section 371(c) of this title before the invention thereof by the applicant for patent.

4. Claims 15, 16, 20 to 24, and 28 are rejected under 35 U.S.C. 102(e) as being anticipated by *Polikaitis et al.*

Regarding independent claims 15 and 28, *Polikaitis et al.* discloses a speech recognition method and system, comprising:

"segmenting a voice signal into words and pauses and converting the words into text" – microprocessor 110 has a speech/noise classifier for determining whether each frame is speech or noise; if the classifier identifies a frame as speech, the classifier assigns the frame an SNflag of 1; if the classifier identifies the frame as noise, the classifier assigns the frame an SNflag of 0; SNflag is a control value used to classify the frames (column 4, lines 31 to 41: Figure 1); a frame classified as speech corresponds to a "word" and a frame classified as noise corresponds to a "pause", where the "word boundary" is the transition between speech and background noise; in voice-input-and-control speech recognition technology, a user may input information, then the technology matches the waveform to a particular word, and provides a signal identifying the particular word (column 1, lines 18 to 38); implicitly, inputting information and providing a signal identifying the word involves "converting the words into text";

"determining an average silence volume during the pauses" – NoiseEnergy is the average energy of all the noise frames as designated by an SNflag equal to 0 (column 5, lines 11 to 23);

"determining an average word volume for the words" – SpeechEnergy is the average energy of all speech frames as designated by an SNflag value equal to 1 (column 5, lines 1 to 10);

"calculating a difference between the average word volume and the average silence volume" – in step 260, microprocessor 110 compares the speech waveform parameters to determine whether the user spoke too softly, Error4; if the ratio ("a difference") of SpeechEnergy to NoiseEnergy is less than a sixth threshold value, Thresh6, then the speech signal is obscured by noise; while any values may be used for Thresh6, Thresh6 is preferably in the range of 6 dB - 24 dB (column 8, lines 46 to 55: Figure 2); the comparison of the ratio of SpeechEnergy to NoiseEnergy is a calculation of a difference between the average word volume and the average silence volume; a ratio represents a "difference" because a larger ratio implies a larger difference and a smaller ratio implies a smaller difference; particularly, sound energies are designated by decibel levels, so that a ratio of sound energies in decibels corresponds to a subtraction of logarithms; a decibel (dB) is defined as "a unit for expressing the ratio of two amounts of electric or acoustic signal power equal to 10 times the common logarithm of this ratio" (Merriam-Webster's Dictionary);

"evaluating a word, having a volume distance between the average word volume and the average silence volume is lower than a predetermined threshold, as having been incorrectly recognized" – in step 260, microprocessor 110 compares the speech waveform parameters to determine whether the user spoke too softly, Error4; if the ratio of SpeechEnergy to NoiseEnergy is less than a sixth threshold value, Thresh6, then the speech signal is obscured by noise (column 8, lines 46 to 55: Figure 2); in step 260, if the ratio of SpeechEnergy to NoiseEnergy is greater than or equal to Thresh6, then the method proceeds to step 290; at step 290, microprocessor 110 performs the speech

recognition process on the speech signal for transmission of a speech recognition signal to the communication interface circuitry 115 (column 9, lines 19 to 34: Figure 2); if Control4 is option C, the user is informed in step 280 that the speech recognition output may be incorrect due to Error4 (column 9, lines 13 to 19: Steps 263, 268, and 280: Figure 2); thus, if an error condition exists, then the speech recognition unit screens the input as incorrectly recognized.

Regarding claim 16, *Polikaitis et al.* discloses that, while any values may be used for Thresh6, Thresh6 is preferably in the range of 6 dB - 24 dB (column 8, lines 46 to 55: Figure 2); a decibel (dB) is defined as "a unit for expressing the ratio of two amounts of electric or acoustic signal power equal to 10 times the common logarithm of this ratio" (Merriam-Webster's Dictionary); thus, *Polikaitis et al.* discloses implicitly that SpeechEnergy and NoiseEnergy are also measured in decibels, which are logarithmic units.

Regarding claim 20, *Polikaitis et al.* discloses Thresh6 is set by the manufacturer preferably (column 8, lines 52 to 54); thus Thresh6 is a constant.

Regarding claim 21, *Polikaitis et al.* discloses no speech recognition is performed if the ratio SpeechEnergy/NoiseEnergy is less than Thresh6 (column 8, lines 46 to 55: Figure 2); instead, an error procedure is performed.

Regarding claim 22, *Polikaitis et al.* discloses in step 263, microprocessor 110 informs the user that Error 4 has occurred; microprocessor 110 communicates Error4 information via the communication output mechanism – communication interface

circuitry 115, speaker 135, display 150, and vibrator/buzzer 160; the information may be communicated through a single output device or any combination of output devices (column 8, lines 55 to 62: Figures 1 and 2); Error4 information output through a speaker or display is "a message".

Regarding claims 23 and 24, *Polikaitis et al.* discloses if Control4 is option A, the user is prompted in step 270 to repeat the voice instruction and is prompted to speak louder (column 9, lines 5 to 8: Figure 2); implicitly, speaking louder causes SpeechEnergy ("average word volume") to increase relative to NoiseEnergy ("average silence volume") as an increased signal-to-noise ratio ("so that an adequate distance is achieved").

Claim Rejections - 35 USC § 103

5. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

6. Claims 17 to 19 and 25 to 26 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Polikaitis et al.* in view of *Wu et al.*

Regarding claim 17, *Polikaitis et al.* discloses SpeechEnergy is the average energy of all speech frames as designated by an SNflag value equal to 1, and NoiseEnergy is the average energy of all the noise frames as designated by an SNflag equal to 0, for all frames 1 to M, where M is the total number of frames (column 5, lines

1 to 23). Thus, SpeechEnergy and NoiseEnergy are global average values, and the ratio SpeechEnergy/NoiseEnergy is a global difference of the values in decibels. Also, *Polikaitis et al.* suggests the user may set or change the value of Thresh6 (column 8, lines 52 to 55). However, *Polikaitis et al.* does not expressly disclose adapting threshold Thresh6 on the basis of the global difference, although adaptive thresholds are fairly well known. *Wu et al.* teaches a generally similar speech recognition method for analyzing endpoints in speech with signal-to-noise ratios, where speech recognition is only performed if a predetermined restart threshold level is identified. (Column 9, Line 56 to Column 10, Line 5) *Wu et al.* employs adaptive thresholds, T_s, T_e, T_{sr}, T_{er}, defined in terms of an average background noise level N_{bg}, and average speech energy levels, E_{ls} and E_{le}. (Column 7, Line 25 to Column 9, Line 31: Figures 8, 9(a) and 9(b)) Specifically, *Wu et al.* says the method is advantageous for eliminating errors due to mistaking breathing for actual speech. (Column 9, Line 56 to Column 10, Line 5) It would have been obvious to one having ordinary skill in the art to employ adaptive thresholds defined in term of average speech energy and average noise energy as suggested by *Wu et al.* for the Thresh6 of *Polikaitis et al.* in order to eliminate errors due to mistaking breathing for actual speech.

Regarding claim 18 *Wu et al.* discloses the thresholds are related to the signal-to-noise ratios, defined in terms of differences E_{ls} - N_{bg} and E_{le} - N_{bg} (column 8, lines 24 to 65).

Regarding claim 19, *Wu et al.* discloses general formulae for adaptive thresholds T_{sr} and T_{er}, where the thresholds are diminished by a factor $-c_3 N_{bg}$, and c_3 is a constant to account for conditions of unstable background noise (column 9, lines 20 to 31).

Regarding claim 25, *Polikaitis et al.* discloses SpeechEnergy is the average energy of all speech frames as designated by an SNflag value equal to 1, and NoiseEnergy is the average energy of all the noise frames as designated by an SNflag equal to 0, for all frames 1 to M, where M is the total number of frames (column 5, lines 1 to 23). Thus, SpeechEnergy and NoiseEnergy are global average values, and average noise is not measured for individual pauses, with the result that the difference between average word volume and average silence volume is not measured in terms of immediately preceding or immediately following silence energy values. However, *Wu et al.* teaches a generally similar speech recognition method for analyzing endpoints in speech with signal-to-noise ratios, where speech recognition is only performed if a predetermined restart threshold level is identified. (Column 9, Line 56 to Column 10, Line 5) *Wu et al.* determines an average background noise level N_{bg} on the basis of segments of silence energy defining a reliable island. (Column 7, Lines 25 to 42: Figure 8) Similarly, *Wu et al.* determines average speech energy levels, E_{ls} and E_{le}, on the basis of segments of speech energy defining a reliable island. (Column 7, Line 58 to Column 8, Line 23: Figures 9(a) and 9(b)). *Wu et al.* says the method is advantageous for eliminating errors due to mistaking breathing for actual speech. (Column 9, Line 56 to Column 10, Line 5) It would have been obvious to one having ordinary skill in the art to determine a difference between average speech energy and average noise energy in

terms of immediately preceding or immediately following pauses as suggested by *Wu et al.* instead of the global average speech energy and global average noise energy of *Polikaitis et al.* for the purpose of eliminating errors due to mistaking breathing for actual speech.

Regarding claim 26, *Polikaitis et al.* discloses SpeechEnergy is the average energy of all speech frames, and NoiseEnergy is the average energy of all the noise frames, for all frames 1 to M, where M is the total number of frames (column 5, lines 1 to 23). *Polikaitis et al.* discloses NoiseEnergy is a global average value, but omits defining the average silence on the basis of a plurality of successive pauses. *Wu et al.* determines an average background noise level, N_{bg}, on the basis of segments of silence energy defining a reliable island, and similarly, determines average speech energy levels, E_{ls} and E_{le}, on the basis of segments of speech energy defining a reliable island. (Column 7, Line 25 to Column 8, Line 23: Figures 8, 9(a), and 9(b)). *Wu et al.* says the method is advantageous for eliminating errors due to mistaking breathing for actual speech. (Column 9, Line 56 to Column 10, Line 5) It would have been obvious to one having ordinary skill in the art to combine the segmental energy averaging method of *Wu et al.* with the global energy averaging method of *Polikaitis et al.* so as to determine the global average silence energy on the basis of a sum of the energies of successive silence segments for the purpose of eliminating errors due to mistaking breathing for actual speech.

7. Claim 27 is rejected under 35 U.S.C. 103(a) as being unpatentable over *Polikaitis et al.* in view of *Wu et al.* as applied to claims 20 to 26 above, and further in view of *Hamasaki et al.*

Polikaitis et al. omits preparing an n-best list on the basis of the difference between the average word volume of individual words, and determining the word to be inserted into the text according to a criterion of the difference between the average word volume and the average silence volume of the individual spoken words. However, *Hamasaki et al.* teaches a similar speech recognition method, where a signal-to-noise ratio is calculated from the logarithm of the average power of a speech segment and the logarithm of the average noise power. (Column 4, Lines 45 to 62: Figure 6) A recognition candidate determiner 14 determines the number of candidates to present in an n-best list varying according to the value of the SN ratio with respect to a threshold x_p. (Column 3, Line 21 to Column 4, Line 24; Column 6, Line 43 to Column 7, Line 33: Figures 3 and 4) Implicitly, the highest scoring word is inserted into the text. *Hamasaki et al.* says the speech recognition method has the advantage of improving a recognition rate by including words in an n-best list that might be eliminated from the list due to a low signal-to-noise ratio. (Column 2, Lines 5 to 49) It would have been obvious to one having ordinary skill in the art to include the speech recognition method of presenting the number of word candidates in an n-best list depending on the value of the signal-to-noise ratio as suggested by *Hamasaki et al.* in the related speech recognition method of

Polikaitis et al. for the purpose of improving recognition accuracy in the presence of noise.

Response to Arguments

8. Applicants' arguments filed 26 November 2003 have been fully considered but they are not persuasive.

Applicants argue the claimed invention, as amended, performs speech recognition, which yields words, pauses and boundaries between pauses and words. Applicants say the average word volume and the average pause volume is then determined based on the recognition result. Applicants state the result of the recognition is corrected such that a word, whose volume distance between the average word volume and the average silence volume is lower than a predetermined threshold, is evaluated as having been incorrectly recognized. Applicants maintain neither *Polikaitis et al.* nor *Wu et al.* discloses these features. This argument is traversed for the following reasons.

Firstly, it is noted that independent claim 28 is not amended. Thus, any arguments directed to the claimed invention, as amended, do not apply to independent claim 28. Applicants have not specifically identified any features of independent claim 28 which are not anticipated by *Polikaitis et al.*

Secondly, Applicants' arguments amount to a mere allegation of patentability. Applicant's arguments do not specifically point out how the language of the claims

patentably distinguishes over the references. Applicants recite certain features of the claims, as amended, but do not identify what is lacking from *Polikaitis et al.*

Thirdly, *Polikaitis et al.* discloses all of the limitations of independent claim 15, as amended. *Polikaitis et al.* segments a voice communication including speech, other acoustic communication, and noise, on the basis of speech frames, by classifying with speech/noise flags whether each frame is speech or noise. (Column 3, Lines 63 to 67; Column 4, Lines 23 to 41) Also, *Polikaitis et al.* discusses how voice-input-and-control for speech recognition technology provides a signal identifying a waveform as a particular word or a command. (Column 1, Lines 21 to 38) Implicitly, identifying a word for purposes of inputting information (as compared with voice control) produces a word from voice as converted text. Then, *Polikaitis et al.* calculates the average energy of all speech frames with a speech flag, and the average noise energy of all the noise frames with a noise flag. (Column 5, Lines 1 to 23) The average energy of speech frames and the average noise energy of noise frames correspond, respectively, to the claimed average word volume and average silence volume.

Polikaitis et al. then takes a ratio of the average energy of all speech frames, SpeechEnergy, to the average noise energy of all noise frames, NoiseEnergy, where the energy values are measured in decibels. (Column 8, Lines 46 to 55: Figure 2: Step 260) Those skilled in the art know that taking a ratio of two values in decibels is equivalent to subtracting the two values, as decibels are on a logarithmic scale. As a result, the ratio of *Polikaitis et al.* is equivalent to the claimed step of calculating a difference. *Polikaitis et al.* compares the ratio to a threshold, Thresh6, and if the ratio is

less than the threshold, then the speech signal is obscured by noise, showing the user spoke too softly. (Column 8, Lines 46 to 55: Figure 2: Step 260) If the ratio is less than the threshold, Thresh6, then the user is informed of Error4 information. (Column 8, Lines 55 to 62: Figure 2: Step 263) If Control4 is option C, the user is informed in step 280 that the speech recognition output may be incorrect due to Error4. (Column 9, Lines 13 to 19: Figure 2: Step 280) This corresponds to the step of evaluating a word as incorrectly recognized when the distance is lower than a threshold, from independent claim 15, as amended. Thus, *Polikaitis et al.* anticipates independent claim 15, as amended.

Therefore, the rejections of claims 15, 16, 20 to 24, and 28 under 35 U.S.C. 102(e) as being anticipated by *Polikaitis et al.*, of claims 17 to 19 and 25 to 26 under 35 U.S.C. 103(a) as being unpatentable over *Polikaitis et al.* in view of *Wu et al.*, and of claim 27 under 35 U.S.C. 103(a) as being unpatentable over *Polikaitis et al.* in view of *Wu et al.*, and further in view of *Hamasaki et al.*, are proper.

Conclusion

9. **THIS ACTION IS MADE FINAL.** Applicants are reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the

shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Martin Lerner whose telephone number is (703) 308-9064. The examiner can normally be reached on 8:30 AM to 6:00 PM Monday to Thursday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (703) 305-9645. The fax phone numbers for the organization where this application or proceeding is assigned are (703) 872-9314 for regular communications and (703) 872-9315 for After Final communications.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is (703) 305-4700.

RICHEMOND DORVIL
SUPERVISORY PATENT EXAMINER

ml
March 5, 2004